APPLICATION FOR UNITED STATES LETTERS PATENT

INVENTORS: Jin-Soo LEE and Heon-Jun KIM

TITLE: MULTIMEDIA RETRIEVAL METHOD USING MULTI-WEIGHTED FEATURE

ATTORNEYS & ADDRESS: FLESHNER & KIM, LLP
P. O. Box 221200
Chantilly, VA 20153-1200

DOCKET NO.: LGE-0017

MULTIMEDIA RETRIEVAL METHOD USING MULTI-WEIGHTED FEATURE

BACKGROUND OF THE INVENTION

1. Field of the Invention

[1] The present invention relates to a multimedia retrieval method using a plurality of descriptors. In particular, the present invention relates to a multimedia retrieval method providing a more advanced retrieval function using an automatically selected optimum weighted data in accordance with the query descriptor, which has a user-desired format.

2. Background of the Related Art

[2] A multimedia retrieval, including an image retrieval, generally uses combined data with colors, textures or shapes. However, since each multimedia data has a different descriptor for retrieving relevant data, a user can enjoy a far advanced retrieval function using an importance of data regarding the descriptor necessary for retrieving the data, if he or she asks a particular query.

[3] One of the typical multimedia retrieval technologies in the related art uses a user interface through which the user can designate the number of descriptors he or she intends to use for retrieving, every time the user does the multimedia retrieval. The user can perform the multimedia retrieval by designating an importance of each descriptor, for example, 50% for the color and 30% for the texture.

[4] Another example of the multimedia retrieval technology in the related art provides feedback data on an image similar to the image the user is looking for and automatically

calculates the weight of the descriptor. However, this method appears to be disadvantageous because the weight is applied only when the user gives feedback for every query.

[5] On the other hand, an attempt recently has been made to standardize the data necessary for retrieving, such as MPEG-7 and so forth. Thus, it has become possible to standardize the weight data on the descriptor and add the standardized data to the multimedia data. In such case, each weight data on the respective data can be applied to a query all the time, regardless of a request of the user, and an advanced retrieval function may be successfully carried out. In this way, a large number of descriptors may be designated in the standardized data since the data designates all kinds of descriptors required within a possible application range. In reality, however, only part of the descriptors can be used for the query and retrieval depending on the application.

[6] For example, the descriptor explaining a video segment can include a variety of semantic data, like the text data, and low-level descriptors like a color histogram. Especially, some low-level descriptors might describe motion data, since video is used for the multimedia in this case. The low-level descriptors may include the descriptors of still images, such as color, texture and shape data because the video image is a collection of still images in a sense.

[7] If the multimedia retrieval is associated with an application for comparing a video segment and other segments, all types of data including motion, color texture, and shape data can be used. However, if the multimedia retrieval is associated with a comparison of the video segment and the still image, the motion data would not be necessary.

[8] Therefore, the descriptors that are actually used in the multimedia retrieval can be different depending on each application. In this case, the relative weight between the descriptors can be altered.

[9] As another example, suppose that the descriptors included in the video segment are color histograms, representative color data, and texture histograms. Further suppose that Application 1 uses all three descriptors while Application 2 uses the color histograms and texture histograms only. Then, in the case of Application 1, the weight on each descriptor of a particular multimedia data can be designated as 0.5, 0.3 and 0.2, respectively, on the condition that all three descriptors (color histograms, representative color data, and texture histograms) are used. Meanwhile, in the case of Application 2, since the color histograms were not used for the same data, the relative weights 0.3 and 0.2 on the remaining two descriptors, i.e., representative color data and the texture histograms are converted and designated to 0.6 and 0.4, respectively.

[10] However, the method described above is not that practical in reality because it is appropriate only when all of the descriptors in the video segment are unrelated and orthogonal to one another. Unfortunately, this is not true in many cases.

[11] That is, unlike the texture histograms, the color histograms and representative color data have a close relationship to each other in terms of color. Therefore, when all three descriptors (color histograms, representative color data and texture histograms) are used as in Application 1, which has two kinds of data on colors (color histograms and representative color

data) and one kind of texture data, the color data can be used practically more than expected, compared with Application 2 where only one kind of color data is used.

[12] Suppose that each weight of those three descriptors is 0.5, 0.3, and 0.2, respectively. To use all of three descriptors is similar to using color relevant data (2) and texture relevant data (1) at a ratio of 0.8: 0.2 (color: texture).

[13] As described in Application 2, if the representative color data and the texture histograms are combined at the ratio of 0.6: 0.4 in terms of the importance, less color data is used compared to the case of using all three descriptors (color: texture = 0.8: 0.2).

[14] For the reason above, it is better to use all three descriptors than to use two types of descriptors, such as representative color data and texture histograms, for increasing the importance of the representative color data. Thus, in case of using a plurality of combined descriptors, it is possible to apply optimized weights with different ratios from one another according to the combination of each descriptor.

[15] Especially, the weight data in the multimedia data, where different combinations of descriptors are available, should have appropriate and separate weight data in accordance with every possible combination of descriptors, thereby obtaining the results from a highly efficient retrieval.

[16]. Going back to the example again, when the multimedia retrieval is carried out based on the combination of the representative color data and the texture histograms, among the color histograms, representative color data, and texture histograms, it is considered that the weight of the combination lies on the color and the texture and the weight of the color data is

set to be relatively higher, i.e., 0.8: 0.2 or 0.7: 0.3. Thus, it is important to have appropriate weight data for each combination of descriptors in order to obtain a more advanced retrieval function.

[17] It is also true that even the identical multimedia data can bring different retrieval results in accordance with an intention of a query. For instance, if the query intends to find an image with a boat, the retrieval might search for any image that has a boat and not necessarily ocean scenery. On the other hand, if the query intends to find an image with ocean scenery, it might end up with any kind of ocean scenery without a boat or the combination of the two is also possible.

[18] Similarly, if the same data is retrieved with a particular query, the result can be varied depending on the kinds of the query (intention, viewpoint). However, if the viewpoint of the query is manifested by different weights, an intended result can be obtained.

[19] Therefore, the multimedia data should have a plurality of weight feature for obtaining an appropriate retrieval result corresponding to a query. Also, the multimedia data should provide a method for automatically selecting the appropriate weight for the viewpoint of the query by figuring out what the user wished to query. The multimedia data should also provide a method for extracting a plurality of weights for the multimedia data.

SUMMARY OF THE INVENTION

[20] An object of the invention is to solve at least the above problems and/or disadvantages and to provide at least the advantages described hereinafter.

[21] It is, therefore, an object of the present invention to provide a multimedia retrieval method using a multi-weighted feature in order to obtain a more advanced retrieval function.

[22] It is another object of the present invention to provide a multimedia retrieval media for retrieving a multimedia object.

[23] To achieve the above objects, there is provided a multimedia method using a plurality of descriptors for retrieving an object. The method includes studying of a descriptor weight in accordance with a combination of the descriptors, among the plurality of descriptors, for a query and adding the weight to the multimedia descriptor; and retrieving, based on the selected weight corresponding to the combination of descriptors, in accordance with the combination of descriptors for the query at the time of the multimedia retrieval.

[24] Here, the studying of the weight according to the combination of descriptors is made possible through the retrieval result from retrieving images using the combination of the corresponding descriptors or feedback from the user regarding a similar object in connection with group data on any similar objects defined beforehand.

[25] Among other descriptors included in the combination of descriptors above, as the descriptor increases a similarity between similar objects a higher weight is obtained.

[26] Retrieving a selected weight corresponding to the combination of descriptors in accordance with the combination of descriptors for the query, during the multimedia retrieval, further includes measuring a similarity on a similar object the user gave feedback after selecting from the group data on similar objects or the retrieval result, using every weight included in the

multimedia descriptors; and retrieving based on the selected weight outputting the highest similarity among other similarities measured.

[27] In the step of retrieving based on the corresponding weight to the combination of descriptors used in the query during the multimedia retrieval, when the user selects a particular descriptor for the retrieval, only the weight studied of the descriptors selected from the plurality of weights included in the multimedia descriptors is used for the retrieval.

[28] Again, in the step of retrieving based on the corresponding weight to the combination of descriptors used in the query during the multimedia retrieval, when the user designates a query object and a retrieval object, only the weight studied of the predesignated descriptors in accordance with the kinds of retrieval object and query object is used for the retrieval.

[29] In addition, for the purpose of retrieving a multimedia object using a plurality of descriptors, the multimedia retrieval method of the present invention includes a descriptor weight studied according to each viewpoint of a query for the retrieval into the multimedia descriptors; and carrying out the retrieval by selecting the corresponding weight to the viewpoint of the query from the descriptor weights included in the multimedia descriptors.

[30] In addition, the multimedia retrieval medium of the present invention has a plurality of descriptors for retrieving the multimedia object; and data features including optimum weight data in accordance with every combination of the descriptors especially for the query among the plurality of descriptors.

[31] Here, the descriptors are combined differently based on the viewpoints of each query and the weights are differentiated depending on the viewpoints of each query also.

[32] Additionally, the objects of the invention may be achieved in whole or in part by a multimedia retrieval method. The method includes determining a descriptor weight for each of a plurality of descriptors used in a first combination to form an multimedia descriptor query and retrieving a group of first multimedia objects based on the descriptor weights corresponding to the first combination of descriptors.

[33] Additional advantages, objects, and features of the invention will be set forth in part in the description which follows and in part will become apparent to those having ordinary skill in the art upon examination of the following or may be learned from practice of the invention. The objects and advantages of the invention may be realized and attained as particularly pointed out in the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[34] The invention will be described in detail with reference to the following drawings in which like reference numerals refer to like elements wherein:

[35] Fig. 1 illustrates an extraction algorithm of weight among other studying methods of a weight based on a combination of descriptors, in accordance with a preferred embodiment of the present invention;

[36] Fig. 2 illustrates a modified sigmoid function in order to obtain a reliability which is applied to the present invention;

[37] Fig. 3 illustrates a diagram explaining application examples using multiweight, which is studied differently based on a combination of descriptors, in accordance with another preferred embodiment of the present invention;

[38] Fig. 4 illustrates a display at the result of a query object and response in accordance with another preferred embodiment of the present invention;

[39] Fig. 5 illustrates an extraction algorithm of weight based on a viewpoint of a query in accordance with another preferred embodiment of the present invention;

[40] Fig. 6 illustrates a weight feature, including description data, on a viewpoint of a query in accordance with another preferred embodiment of the present invention;

[41] Figs. 7a and 7b illustrate diagrams showing an example of a query image and a query viewpoint in accordance with another preferred embodiment of the present invention; and

[42] Fig. 8 illustrates a diagram showing a weight feature, including data on descriptors, in accordance with another preferred embodiment of the present invention.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[43] For a better understanding of a multimedia retrieval and a weight studying method, the present invention is divided into four categories: (1) a studying method of multiweight according to a combination of descriptors; (2) an application method of multiweight according to a combination of descriptors; (3) a studying method of multiweight according to a viewpoint of a query; and (4) an application method of multiweight according to a viewpoint of a query.

[44] **1. A Studying Method of Multiweight According to Combination of Descriptors**

[45] Suppose that a set of the entire descriptors for the multimedia retrieval is Set A= [C1, C2, C3, …. , CN] and a set of the descriptors used for the combination of current descriptors is Set B= [Ci1, Ci2, Ci3, …. , CM] $\subseteq$ Set A. An algorithm for extracting weights for the descriptor set B is explained in Fig. 1.

[46] First of all, each descriptor of Set B, [Ci1, Ci2, Ci3, …. , CM], is set to have an equivalent importance and the similarities thereof are measured and arranged (S101).

[47] Any type of error in the arranged retrieval result is retrieved and the number of errors is designated as E (S102). In other words, if the number of correct images is k, the error means a lower-ranking image than a high-ranking k regardless of the fact that the error is a correct image as well. The number of errors is set at E.

[48] The correct image data can be obtained through two methods: the user gives feedback on the correct image data in a first retrieval result obtained from step 101; and the images in the same class are collected from a database and grouped beforehand. Later, the correct image data is found using the sample group.

[49] Next, if the error, E (or E/k, wherein E is the number of the error images, E/k is a degree of error from a point of accuracy in percentage), is equal to or below a specific critical value, Th1, the studying method is completed (S103). If the error is not below the critical value described above, the number of images necessary for the feedback is decided using the error E, using the equation n=2 x E (S104).

[50] The feedback is then given as many as the number of the feedback images decided before (S105). Here, n number of the error pages having the most errors are arbitrarily selected, among the error pages used in step 102, and are designated as feedback images automatically. The lower the image ranks, the more error exists. A weight for each descriptor is updated based on the feedback given (S106).

[51] Using the updated weights, the image is retrieved again and arranged (S107). Any error shown in the list of the retrieval result is designated as E' (S108). If the error, E' (or E'/k), is equal to or below the specific critical value, Th1, the studying method is completed, but if not, it proceeds to the next step (S109).

[52] The degree of decrease in the error is examined to determine whether it exceeds a specific critical value or not (S110). That is, if E-E' is larger than the specific critical value (Th2) (or E x Th2' > E), the number of feedback images, n, in the following step is determined using the current error E', such as, n = E' x 2 (S111). If E-E' is not larger than the specific critical value (Th2), the number of feedback images is determined as a smaller number than the previous feedback number, such as, n = n x b, where 0<b<1 (S112).

[53] In case the number of images necessary for the feedback is even smaller than the minimum (it is 2 in the present invention, 2 = a similar image 1 + a dissimilar image 1), the number of the feedback images is designated as 2.

[54] Lastly, the current E' is re-designated as E and the studying method goes back to step 105 (S113).

[55] In this way, the weight according to the combination of descriptors, namely the weight for Set B, was extracted and the weight was updated using the feedback images in step 106. The method for updating the weight is explained below.

[56] A new weight (New_W) to be updated is first calculated:

[57] New_W = (Reliability x Old_W + Cur_W) / (Reliability + 1) (1)

[58] Wherein, Cur_W is a relative weight calculated based on the feedback given currently. Old_W is a weight before the renewal. Reliability is a value of reliability of the current weight.

[59] Evidently, the reliability indicates the extent to which the weight is reliable. And, the new reliability (New_R) (new Reliability to be updated) can be calculated as follows:

[60] New_R = old_R (1 + Increase R) + a (2)

[61] Increase R = f (# feedback) x (Precision (t) – Precision (t –1)), (3)

[62] wherein # feedback is the number provided by feedback, such as the number of relevant images selected by a user in one stage. Old_R is the previous reliability and serves as the value of Reliability in equation (1). Precision (t) is the precision of a result at time t and may be determined by a user's feedback. Preferably, the precision may be calculated as:

[63] Precision = (the number of relevant images in a rank N)/N, where N has a temporarily defined value.

[64] Moreover, if f(# feedback) is small, equation (3) returns an approximate value of 0 and if f gets larger, equation (3) returns an increment. A plot of equation (3) is shown in Fig. 2 and a modified sigmoid function can be used in this case to represent the equation.

[65] As explained before, the relative weight based on the currently given feedback (Cur_W) can be calculated as follows:

[66] Cur_W = a Sim (RI, FI), supposing that FI is a relevant image and

[67] Cur_W = a Dist (RI, FI), supposing that FI is an irrelevant image,

[68] wherein FI is a feedback image; RI is a reference image; W is a type weight (Wk), an element weight (We) or a position weight (Wp); and a is a normalization coefficient for Wk, We and Wp. The relevancy of the retrieved images is preferably determined by the user's feedback.

[69] In addition, Sim (FI, RI) indicates the similarity between the reference image (RI) and the feedback image (FI), in the case of using the descriptors, k, e and p. On the other hand, Dist (FI, RI) indicates the dissimilarity between the reference image (RI) and the feedback image (FI), in the case of using the descriptors, k, e and p.

[70] Therefore, when the retrieval is carried out using the combination of descriptors in Set B, the corresponding weight to the image is successfully studied as described above.

[71] Similarly, if there exists another combination of descriptors, Set C, appropriate weights can be retrieved and studied using the descriptors included in Set C. At this time, the dimension of a weight for each descriptor is the same as the number of constituents in the Set.

[72] **2. Application Method of Multiweight According to Combination of Descriptors**

[73] As explained before, differently studied multiweight according to each combination of descriptors can be applied to other cases. To begin with, the user can select a

descriptor for the multimedia retrieval using a user interface, as shown in Fig. 3. Particularly in Fig. 3, color histograms, representative color data, and motion histograms are selected, for example, out of five descriptors (color histograms, texture histograms, representative color data, motion histograms, and shape data). Once the user selects descriptors necessary for the multimedia retrieval using the user interface, the weights thereof are automatically selected for the retrieval, based on the method explained with reference to Fig. 1 and according to the combination of descriptors selected.

[74] A method of selecting a weight out of a plurality of weights, based on the study on the selected descriptors only, is now explained below.

[75] The weight feature includes each weight and data showing which weight goes with which descriptor. Based on the aforementioned data, it becomes possible to find out which weight of a descriptor is included in the current weight and the descriptors, represented by the weights included, select the weights corresponding to the descriptors selected.

[76] Another method of selecting multiweight is that the user selects a similar object to what he or she intended to find from the early retrieval result. As for the similar object selected, the similarity is measured using a respective weight that is studied according to the combination of descriptors and the weight outputting the result, based on the highest similarity, is selected.

[77] If the multimedia object already contains data on an example of the objects belonging to the similar class, the retrieval method displays the identical class object with the query object (in this case, it is an image), as shown in Fig. 4, and helps the user to select the

object he or she intended to find. In other words, the method displays the identical class image with the query image, as the similar image, and makes it possible for the user to select the image he or she wanted.

[78] The similar image surrounded by a bold rectangle in Fig. 4 is the object the user selected as the similar image. Once the object is selected, the similarity between the similar image selected by the user and the query image is measured using each studied weight, in accordance with the combination of descriptors, and the weight outputting the highest similarity is selected.

[79] Also, the weight can be selected by different kinds of queries that are designated in advance. For example, the queries can be divided into several kinds, e.g., video and video retrieval, or video and still image retrieval. The descriptors for each case are determined and the corresponding weights are designated in order to select the appropriate weight, in accordance with the kind of the query of the user.

[80] The kind of query the user asked can be determined once a query object and a retrieval object are determined. For instance, in the case that the kind of the query is a video segment and the retrieval object is an image database, the retrieval identifies video and still images.

[81] **3. Studying Method of Multiweight According to Viewpoint of Query**

[82] Fig. 5 illustrates an example explaining the algorithm for extracting the weight in accordance with the viewpoint of a particular query. First of all, the descriptors in possession are set at an equivalent importance and the similarity thereof is measured and arranged later

(S500). Next, the user gives feedback on the similar image, corresponding to the viewpoint of the current query, in the early retrieval result (S501). In the following steps, the similar image from the feedback is considered as the correct answer and the studying is proceeded from there.

[83] Especially in steps of 500 and 501, the data on the correct answer according to the viewpoint of the particular query was retrieved based on the feedback given by the user from the early retrieval result. If every object contains the data on image examples belonging to the same class, instead of carrying out the early retrieval as in step 500, the image examples belonging to the same class are displayed first and the user can select the similar image among the displayed images.

[84] Beginning with step S502, the same procedure as that of Fig. 1, beginning with step 102, is repeated here. That is, the number of errors, E, is designated by retrieving any error in the arranged retrieval result (S502). If the error E (or E/k) is equal to or below the specific critical value, Th1, the studying method is completed. If not, the number of images necessary for the feedback (n = 2 x E) is decided using the error E (S503 and S504). The feedback is given as many as the number of images, n, in accordance with the method described above (S505).

[85] After that, a weight of each descriptor is updated using the feedback given (S506) and, based on the updated weight, the image is retrieved again and arranged (S507). Another error, E', is designated after retrieving the error shown in the list of the retrieval result (S508). Here, if E' (or E'/k) is equal to or below the specific critical value, Th1, the method is completed, but if not, the next step proceeds (S509).

[86] The degree of decrease in the error is examined to determine whether or not it exceeds the specific critical value, Th2, (S510). Based on the result obtained here, the number of images, n, necessary for the feedback is determined (n = E' x 2) using the current error E' (S511). Otherwise, the number of images for the feedback can be decided to be smaller than the number of the previous feedback (n = n x b, 0<b<1) in the next step S512. After re-designating the current error (E') as E, step S505 is repeated (S513).

[87] **4. Application Method of Multiweight According to Viewpoint of Query**

[88] As explained before, differently studied multiweight according to each combination of descriptors can be applied to other cases. In the beginning, the user selects the similar object to the object he or she intended to find in the early retrieval result. As for the similar object selected, the similarity is measured using the respective weight that is studied according to the combination of descriptors and the weight outputting the result based on the highest similarity is selected.

[89] If the multimedia object already contains data on an example of the objects belonging to the similar class, the retrieval method displays the identical class object with the query object, as shown in Fig. 4, and helps the user to select the object he or she intended to find. Once the object is selected, the similarity is measured using each weight and the weight appropriate for the corresponding viewpoint of the query outputting the result of the highest similarity is selected.

[90] Fig. 6 shows an example of the weight feature, which contains a weight scheme for the viewpoint of the query. The weight scheme 600 is comprised of a descriptor ID 602, for

describing weights 601, the corresponding weight value 603, and scheme of the viewpoint for the query 604. If the viewpoint of the corresponding query in the weight feature is described in text, as shown in Fig. 6, this also can be applied.

[91] In other words, the viewpoints of the queries of existing weights (especially the viewpoints of the queries described in text) are arranged, as shown in Fig. 7b, to be selected by the user. Particularly, Fig. 7a illustrates an image of landscape including a house, where the viewpoint of the query can be things like a house, a blue sky, or a field.

[92] On the other hand, Fig. 8 shows a weight feature containing data on the combination of descriptors. The weight scheme 800 is comprised of a descriptor ID 802, for describing weights 801, the corresponding weight value 803, and a list of associated descriptors 604.

[93] Using the weight feature containing data on the combination of descriptors, it becomes easier to select an appropriate weight through an immediate use of the data.

[94] Until now, a multi-weight feature according to the viewpoint of the query and the multi-weighted feature according to the combination of descriptors have been explained.

[95] Therefore, using the multi-weighted feature according to the combination of descriptors and the multi-weighted feature according to the viewpoint of the query, the corresponding weight can be automatically selected and used despite the fact that each weight feature does not necessary have data on the combination of the descriptors or the viewpoint of the query itself.

[96] However, if each weight feature contains data on the combination of descriptors currently used for the weight, as shown in Fig. 8, or the viewpoint of the query is described in each weight feature, as shown in Fig. 6, it is much easier to select an appropriate weight using the data immediately.

[97] In such case, although a more easy application is possible, the same can be regarded defective in terms of the size of data since additional data are required.

[98] As explained hitherto, the multimedia retrieval method of the present invention enables a far advanced retrieval function by providing each object with the optimum descriptor weight that is appropriate for a variety of applications. In addition, the present invention enables the user to carry out different retrievals according to different viewpoints even on the same object, by extracting an optimum weight for the viewpoint of the query the user asks, thereby making it come true of the user-centered retrieval.

[99] The foregoing embodiments and advantages are merely exemplary and are not to be construed as limiting the present invention. The present teaching can be readily applied to other types of apparatuses. The description of the present invention is intended to be illustrative, and not to limit the scope of the claims. Many alternatives, modifications, and variations will be apparent to those skilled in the art. In the claims, means-plus-function clauses are intended to cover the structures described herein as performing the recited function and not only structural equivalents but also equivalent structures.